



US Department of Transportation
Office of the Chief Information Officer
Model Data Inventory Approach

November 26, 2013



Background

The purpose of this document is to provide a model plan for conducting data inventory efforts required under OMB Memorandum M-13-13, *Open Data Policy – Managing Information as an Asset*.¹ Your data inventory is complete when you have accounted for all datasets that are:

- collected under an OMB-approved Information Collection Request
- referenced in a Records Disposition Schedules (RDS) and Electronic Information Systems (EIS) list
- covered by a Privacy Act System of Records Notice (SORN)
- associated with an information technology (IT) investment reported on your Exhibit 53
- associated with your agency Data Reference Model (DRM) and/or Logical Data Model (LDM), if applicable

You will register all datasets in the DOT metadata registry at <http://data.nhtsa.gov>. If your Component² operates its own registry, contact the DOT Chief Privacy & Information Asset Officer, Claire Barrett (Claire.Barrett@dot.gov) to have your registry evaluated for compliance with OMB requirements.

Model Data Inventory Approach

Forming the Team

To successfully complete your data inventory, you will want to leverage your existing governance structures to form an integrated team from a number of different information management disciplines including but not limited to:

- Information Collection Officer – this individual will have access to the inventory of collections and forms you use to collect information from people and organizations outside the DOT. In addition, this individual will have information about whether information is collected electronically.
- Records Officer – this individual will have access to your organization’s records schedules, which will cover all the types of information you manage, whether collected from outside DOT or generated through administrative activities inside DOT. The schedules may also contain information about protections that are applied to those records. They will also have a list of electronic information systems that contain records.
- Privacy Officer – this individual will have access to your organization’s privacy program documentation, such as privacy threshold analyses, privacy impact analyses, and system of records notices. These documents help identify what systems you have and whether there are privacy protections related to data in your information systems.
- Information Systems Security Officer – this individual will have access to your organization’s list of information systems and supporting documentation regarding the sensitivity of the information residing in those systems.

¹ <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>. Supplemental guidance may be found at - <http://project-open-data.github.io/implementation-guide/>

² The term Component is used throughout the document to refer to all DOT Operating Administrations and Secretarial Offices.

- Enterprise Architect – this individual will have access to information about your systems and the data they contain, and may have access to documentation that describe the contents of those systems in detail.
- Capital Planning and Investment Control Coordinator – this individual will have access to your organization’s list of information technology investments.

You may also wish to include key program personnel, such as data stewards, as they may also have access to key documentation or policies on the releasability of data and information from their systems. Other optional personnel could include your Freedom of Information Act (FOIA) officer, who may have knowledge of the types and frequency of requests for certain kinds of data you hold.

Establishing the Universe

The process of identifying your organization’s data holdings will link a number of different information management program areas. Each program area has its own approach to managing information, which can make one-for-one links between them difficult. Figure 1 below identifies key connections.

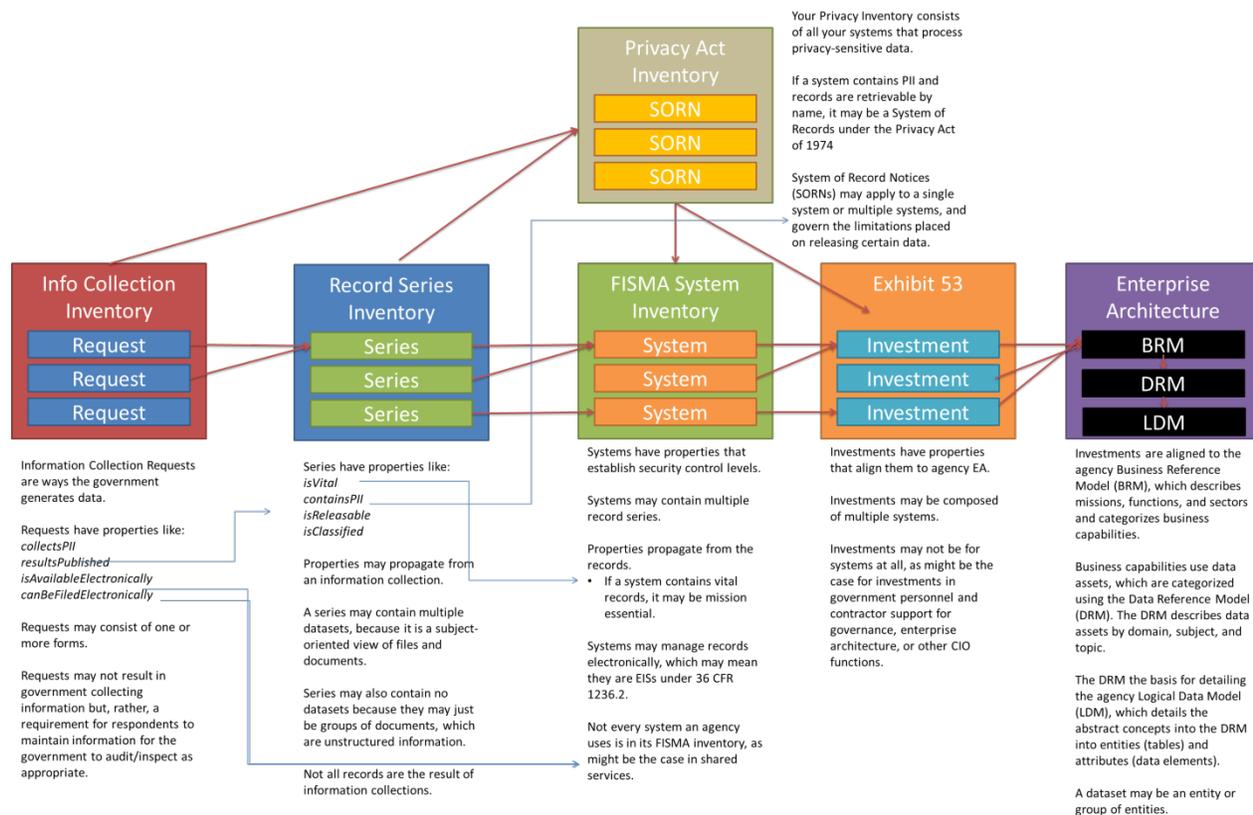


Figure 1 - Leveraging multiple governance disciplines results in a comprehensive view of Component information assets

Your data inventory should begin with an evaluation of your Component information collections. You should review your active information collections and map them to your RDSs, your EIS list, and your Exhibit 53.³ The purpose of this activity is to understand how much information you collect from

³ A listing of your information collections subject to the Paperwork Reduction Act may be found at <http://www.reginfo.gov/public/do/PRAXML>; however it should be noted that this is not an exhaustive list, many agencies have information collections not subject to the PRA.

external sources and how that information is collected – for example, which forms are used to collect the information. As part of the justifications that are submitted to OMB, your organization may indicate whether the results will be published. In general, your information collection justifications may indicate whether indicate the results of information collections will be published or that indicate you are providing access to that information on your Web site. In addition, your information collection justifications may indicate whether they are available electronically and that can be submitted electronically. These collections may be linked to databases and information systems that could provide extracts or open data. Finally, your information collection justifications may indicate whether you are collecting personally identifiable information (PII).

The purpose of linking information collections with your RDSs and EIS inventory is to help understand whether those data are temporary or permanent records and which information system contains the data. Linking your EIS inventory to your Exhibit 53 will help you verify that you have considered all of your technology investments.

This activity will help you to understand the data your organization generates through administrative or program management activities – information you would not find solely through your information collection inventory. Your records schedules will help you identify the types of records that are covered, and whether those records are paper, electronic, or both. It is important to understand that your RDS is not, in itself, a data inventory. RDSs are typically generated at the “records series” level, and a record series may contain multiple data sets because it is a subject-oriented view of files and documents. In general, your data inventory efforts are focused on records that are stored electronically, which should mean that they are stored in an EIS. You will map these to your Exhibit 53. The purpose of linking your EIS list to your Exhibit 53 is to ensure that there is a clear link between your IT investments and your systems. A key part of the OMB memorandum on open data is that any substantially modernized information system must plan for making its data open. Establishing this kind of traceability will help you comply with the policy.

Next, you will review your systems inventory⁴ and SORNs⁵ and map these to your Exhibit 53. If your agency has conducted a biennial SORN review, you may have already documented these relationships. The purpose of this activity is to understand your Privacy Act inventory, and the limitations placed on releasing certain data collections. Because our data inventory will include *all* of our data sets, it is important that we keep track of these kinds of protections and restrictions in a central place. A key part of the OMB memorandum is that we must be transparent about why we are not releasing data sets, and linking to SORNs will help us comply with that objective.

Integrating releasability and prioritization assessments into the inventory process will limit the amount of rework necessary during subsequent Open Data activities. The Department’s “Data.gov Interim Identification & Prioritization Process and Guidelines v1.0”⁶ establish a uniform set of criteria including value, quality, and manageability which when taken together, provide a uniform basis for selecting and prioritizing the DOT’s data release.

Inventorying Datasets

Once you have a complete list of information collections, record schedules, and IT investments that contain data, you will proceed with inventorying the datasets contained in those information systems.

⁴ Federal Information Security Management Act (FISMA)-reportable systems inventory

⁵ You can obtain a list of your SORNs at <http://www.dot.gov/individuals/privacy/privacy-act-system-records-notice>

⁶ <http://www.dot.gov/sites/dot.dev/files/docs/identpriorguidelines1.0.pdf>

You do not need to inventory each and every table contained in and report generated by the information system. A dataset is a group of data elements that make sense to group together. Each Component may identify datasets specific to supporting the needs of their respective mission or business functions. Notably, a dataset is a deliberately abstract concept. A given dataset may represent an entire database consisting of multiple distinct subject areas (for example, a database of crash records could contain information about people, vehicles, and the crash itself), or may represent a single subject area (for example, a database containing information about bridges, roads, or tunnels).

As matter of prioritization, it is strongly recommended Components inventory datasets in the following order:

1. Datasets that are already made available to the public through Component Web sites
2. Mission-oriented datasets that may not yet be made available to the public
3. Administrative datasets that may or may not yet be made available to the public

Components may use alternate prioritization approaches so long as that approach results in a complete inventory all datasets no later than September 29, 2014.

Project Plan

A project plan for this approach is outlined below.

Activity	Deadline
Develop Component implementation plan	January 17, 2014
Establish the Universe	February 28, 2014
Inventory public datasets & initial releasability analysis	April 28, 2014
Inventory mission-oriented datasets & initial releasability analysis	June 30, 2014
Inventory administrative datasets & initial releasability analysis	September 29, 2014